DOCUMENT RESUME

ED 130 832 SE 020 882

AUTHOR Rezba, Richard; And Others
TITLE Evaluating Teacher and Program Effectiveness Through Pupil Performance Measures.
PUB DATE Apr 76
NOTE 20p.; Paper presented at the annual meeting of the National Association for Research in Science Teaching (49th, San Francisco, California, April 23-25, 1976); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Educational Research; Effective Teaching; *Evaluation; *Performance Criteria; *Preservice Education; Program Effectiveness; *Science Education; Secondary School Teachers; *Teacher Education
IDENTIFIERS Research Reports

ABSTRACT

This study was designed to: (1) design pupil growth measures for evaluating teacher and program effectiveness; (2) evaluate the effect of training site on preservice teachers' ability to bring about intended outcomes in pupil learning; and (3) determine the extent to which preservice teachers could bring about intended outcomes in pupil learning. Using classification skills and understanding of ecological concepts as pupil growth measures, it was shown that pupils taught by preservice teachers scored higher than students who were not instructed by preservice teachers. Also, students taught by field-based intern teachers scored higher than students taught by university-based interns. (MH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED130832

U S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Forty-Ninth Annual Meeting

NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING

San Francisco, California April 23-24, 1976

EVALUATING TEACHER AND PROGRAM EFFECTIVENESS THROUGH PUPIL PERFORMANCE MEASURES

Richard Rezba
Anton Lahnston
Diane Lapp

School of Education
Boston University
Boston, Massachusetts 02215

BEST COPY AVAILABLE

SE 020 882

The question of whether the curriculum of teacher preparation programs makes any difference in pupil performance is receiving increasing attention. This is due in part to the trend toward competency-based teacher education (Popham, 1973) and to the continuing search for criteria by which to assess the effectiveness of trainees and teacher training programs (Popham, 1971 Turner, 1972).

Teacher education is continually faced with the problem of the assessment of future teachers and teacher preparation programs. Traditionally, the effectiveness of teachers and their training programs has been evaluated on the basis of knowledge preservice teachers have acquired and to some extent on their demonstrated teaching skills. More recently researchers have advocated the use of pupil performance measures (Dunkin and Biddle, 1974; Popham, 1971 1973). Additional support from researchers and professional associations for the position that the ultimate criterion of teachers' competence is their impact upon learners is cited by McNeil and Popham, (1973).

Popham (1973) suggested that an outcomes-focused approach to teacher education rests on the central assumption that success of teachers is measured not by what teachers do, but what happens to pupils as a consequence of what the teacher does. He added, however, that significant ends can only be brought about by employing appropriate means. Thus, outcome focused programs must provide instruction that enhance the teachers' skills in using a wide variety of teaching techniques. Although many process variables (teacher behaviors) have been researched and advocated as being beneficial, the relationships between these process variables and intended pupil growth are inconclusive (Rosenshine and Furst, 1971). While these relationships are not clear, they do provide tentative guidelines for skill development.

Turner (1972) listed six criteria that could be used to assess the effectiveness of teacher-education programs. At the highest level, the two-part criterion Level 1 consists of the observation of the behaviors exhibited by the teacher and the systematic analysis of the level of outcomes achieved by the teacher with the pupils taught over a period of at least two years on a time-sampling basis. Criterion Level 2 is identical to Level 1 except that shorter performance is involved. Level 3 refers to the effects of training on a teacher's behavior under actual classroom conditions. Criterion Levels 4 and 5 are concerned with the effects of teacher training on improvement in pedagogic skills under laboratory or simplified training conditions. Level 6 is concerned with effects of a training program on improvements in teacher knowledges and understandings.

Criterion levels 3 through 6 focus directly on the impact of training on teacher behavior, while Levels 1 and 2 are concerned with both the effects of training programs on teacher behavior and with the effects of teacher behavior on pupil performance.

Turner suggested that the most appropriate criterion level for accountability in teacher education is Level 3. He acknowledged, however, that some teacher educators argue that Levels 1 and 2 are the most appropriate levels for assessing the effectiveness of training. He added that the emphasis on pupil change in these two levels equates accountability in teacher education with school accountability. Teacher education does not address itself directly to the modification of pupil behavior and it is doubtful that measures of school accountability are appropriate measures of the effectiveness of teacher-education programs. Harste (1973) argued, however, that at some point teacher educators must be held accountable for the normative or long-range growth of pupils affected by teachers trained in teacher education programs.

Few educators, if any, would rely solely on pupil growth measures to evaluate either teacher or teacher education programs. However, performance tests of teaching proficiency based on pupil growth measures can contribute to a body of knowledge on the effectiveness of individuals and programs. (Cooper, 1974).

Two assumptions proposed by Lucio and McNeil (1969) as the ultimate criteria for teaching effectiveness include: 1. learning is evidenced by a change in behavior and 2. teaching is successful when intended changes sought in learners actually occur. Thus, pupil performance criteria is thought to be less inferential and more predictive than teacher-pupil process (teaching behaviors) or teacher presage (knowledge or background) criteria (Harste, 1973; Justiz 1969; McNeil and Popham, 1973; Schalock, 1971, Smithman and Lucio, 1974), however, while this position is held by many, the practical use of pupil performance measures has rarely been made (Cooper, 1974; Rosenshine and Furst, 1971).

PURPOSE

The purpose of this paper is to report the development, use and results of pupil performance measures (called teaching performance tests by Popham, 1973, and McNeil and Popham, 1973). Measures described include one in progress (communication skills) and three completed: 1. classification skills; 2. ecological concepts; and 3. observing and inferring skills. Specifically, the objectives of the study were to:

1. Design pupil growth measures for evaluating teacher and program effectiveness.
2. Evaluate the effect of training site on preservice teachers' ability to bring about intended outcomes in pupil learning.
3. Determine the extent to which preservice teachers could bring about intended outcomes in pupil learning.

Since 1973, pupil performance measures have been included as part of a multi-dimensional research design for evaluating Boston University's field-based, junior year block methods program. The research design included measurements made on objectives in four broad domains: affective, cognitive performance (teaching behaviors), and consequence (pupil growth). The use of pupil performance measures in the consequence domain is presented in the following sections.

DEVELOPMENT

Prior to the implementation of each of the four pupil performance measures used since 1973 it was necessary to select appropriate topics, derive objectives, gather relevant teacher resources, and construct pre- and posttest items. All proposed topics were subjected to stringent requirements. The first requirement, related to content, was that each topic had to have learning potential for a wide range of elementary children; it had to be sufficiently interesting so that children would want to participate; and sufficiently important so that cooperating teachers would support its inclusion into the curriculum. Furthermore, to reduce the likelihood of previous student experience with the topic, it could not be a regular part of the schools' curriculum. But neither could it be so obscure that cooperating teachers would reject it nor be totally alien to the experience of the junior interns. The second requirement to be met was that of time; selected topics had to be of short duration. The first measure used, classification, was selected for its potential to be accomplished within a short period of time (approximately 45 minutes). The other three measures used were chosen because they could be accomplished by a series of short lessons over a period of a few weeks. Although these measures required additional teaching time, deemed desirable by the staff, they did not require an excessive amount of time

that would impose on the cooperating teachers' curricular responsibilities.

For each of the four measures, the following developmental procedures were conducted. Proposed topics were reviewed and a tentative choice meeting the required criteria was selected. Based on this topic, performance objectives were selected or derived. After extensive discussion with staff and colleagues, the objectives were revised as needed and parallel pre- and posttest items based on the objectives were developed. Test items were then piloted with elementary students from each grade level, K through 6, on a one to one basis in schools not involved in the study. The testing process was taped onto cassettes so that they could later be re-played by the staff for possible test revision.

The purpose of the piloting was to: 1. test the sequence of objectives; 2. check the length of the test relative to students' attention span; and 3. gather some evidence of the validity of test questions and associated materials (manipulatives or pictures) as test media. Due to the range of reading levels of the children to be instructed, the tests necessarily had to be oral and pictorial. Piloting always led to the revision of some test items to insure clarity. Cooperating teachers in the schools where the interns worked reviewed the selected topic and made recommendations on both the content and procedures for implementing the tasks in their classrooms.

Sources of ideas and information related to the chosen topic were gathered for the preservice teachers (junior interns) to use if they desired. Children's tradebooks, associated materials and supplies were also assembled and made available to the interns.

During each of the semesters in which a pupil performance measure was used, related workshops were conducted for the junior interns during regular class time or during special class sessions. This was for the purpose of

1. introducing the rationale for the task; 2. explaining the timetable and procedures to be followed and 3. describing the diagnostic instruments and the administration procedures. At that time the nature of the posttest as a parallel form of the pretest was explained although the exact form of the posttest was not shown to them until they were ready to posttest their students. The workshops also provided the opportunity to instruct the interns in the content background relevant to the topic and necessary for them to design instruction for their students.

Each of the teaching tasks involving pupil performance measures required the interns to perform the following activities:

1. diagnose the children's knowledge of the topic,
2. prepare written instructional plans to aid students in the attainment of the objectives,
3. teach a group of 4-5 pupils using their plans.
4. evaluate student mastery of the objectives.

The first pupil performance measure on classification skills, was used to test the effect of training site on preservice teachers' ability to bring about specified learning outcomes in pupils (objective 2). This was accomplished through the pre- and posttesting of 202 pupils in thirty classrooms of the six school sites. An analysis of covariance design was generated to test differences between pupil groups. Pretest scores were used as a covariate to adjust the dependent variate, performance scores on the posttest. Fifty-nine junior interns were randomly assigned to two experimental groups, Group A (in a totally field-based program) and group B (in a university-based program).

Field experience for the interns took place in six Boston area (urban) public schools. A grade range of K-8 was included with a total of 30 teachers voluntarily participating. The educational level of the teachers and their years of teaching experience were similar. All schools consisted of a

socially and economically heterogenous pupil population.

The treatment for the juniors consisted of one full semester of instruction in the content areas of the block methods program (reading-language arts, science and social studies). The field-based interns (group A) received all instruction in a public school classroom. Instruction in the content areas was interwoven with experience in the classroom four days a week. For the university-based, field-oriented interns (group B), instruction in the same content areas for the same amount of time by the same instructors took place at the university on one day with a full-day of supervised field experience on another day. The "classification" pupil performance measure consisted of seven objectives (e.g. Construct and demonstrate the use of a binary classification of objects according to variations in a single characteristic that the student has chosen). Seven tasks were designed to measure the attainment of the objectives (e.g. Tell the child "put the objects into two groups so that the objects in each group are the same in some way"). More detailed descriptions of the experimental groups and the "classification" pupil performance measure are available elsewhere (Johnston et al, 1974; Rezba et al, 1974).

In another phase of this research, the ability of preservice teachers to bring about intended outcomes in their students objectives was further investigated using a pupil performance measure based on ecological concepts. A new sample of forty-five juniors in a field-based program was given a specified set of eight objectives (e.g. Given a diagram of selected plants and animals, the student should be able to construct a three-part food chain). Fifteen tasks and subtasks were designed to measure the attainment of the objectives (e.g. "What might happen if these living things were all living in a terrarium?"; "Show what happens by drawing arrows in the picture").

The task of the interns consisted of designing and implementing instruction on ecological concepts adapted from the SCIS Populations Unit. For this task, it was necessary to provide the interns with extensive background information and materials as well as equipment and living organisms. The measurement of the ecological task was accomplished through the pre- and posttesting of 171 K-8 pupils in 32 classrooms in six school sites. Pre- and posttest means were computed as well as three measures of central tendency of pupil gain scores. Detailed information on the "ecological concepts" pupil performance measure are available elsewhere (Lapp et al, 1975).

A third teaching task using a pupil performance measure and a different population of interns was based on "observing and inferring skills". Twenty-three junior interns received instruction in the basic process skills of observation and inference. They were given the task of designing instruction based on pretest results of their students on eight specified objectives (e.g. Distinguish between statements that are observations and those that are inferences). Eight parallel tasks for both the pre- and posttests were generated (e.g. Give the child a sealed milk carton. Say: "I am going to read three statements about the box. Tell me whether each statement is an observation or an inference"). Evaluation of the entry and mastery levels on the eight observation and inference objectives was accomplished through the pre- and posttesting of 89 K-6 pupils in 18 classrooms of five school sites. Pre- and posttest means were computed as well as three measures of central tendency of pupil gain scores.

The fourth teaching task using a pupil performance measure was based on "communication skills" and is currently in progress using 14 classrooms in five school sites and involving 16 junior interns. Communication was chosen as the teaching task theme for the following reasons:

1. It is a broad theme with a range of related topic areas and offers the possibility of a variety of interdisciplinary approaches.

2. Communication, as a topic area, presents a range of concepts and skills suitable to the span of learning levels K-8 represented by the pupil participants.

3. Previous teaching task themes, although interdisciplinary in nature, had been drawn from science materials. For this task, it was decided to choose a theme from social studies materials but as in the past one that offered the possibilities for interdisciplinary approaches.

A list of seven objectives were drawn from the topic area of communication (e.g. The student will identify examples of non-verbal communication). For each of the objectives, parallel test items were designed (e.g. Point to the picture that shows a person getting an idea without the help of words). More detailed information on both the "observing and inferring skills" and the "communication skills" pupil performance measures are also available (Rapp et al, 1976).

RESULTS

Classification Task

(Insert Tables 1 and 2 about here)

In Table 1, the mean for the pretest indicates the number of objectives that the pupils successfully completed prior to instruction by the interns while the posttest mean indicates the number of objectives that were successfully completed after instruction. The adjusted mean reflects the use of the pretest as a covariate adjustor on the posttest scores.

The analysis of covariance summary presented in Table 2 shows a significant difference between pupils instructed by field-based interns and university based, field-oriented interns as measured by the consequence posttest ($F=17.187$, d.f. 1, 199, $p < .01$).

(Insert Figure 1)

Figure 1 shows the difference between pupils taught by the field-based interns (A) and pupils taught by the university-based, field-oriented interns (B). The pretest mean score of pupils in group A ($\bar{x}$ = 3.691) was lower than that of pupils in group B ($\bar{x}$ = 4.182). However, the posttest mean of pupils in group A was higher ($\bar{x}$=5.642) than that of pupils in group B ($\bar{x}$ = 5.003) The highest possible score was seven (7).

The significantly greater growth in classification skills of pupils instructed by the field-based interns may be attributable to the increased time that the interns had to practice classroom teaching skills and to develop interpersonal relationships with pupils.

Ecological Task

The results of the ecological concepts task revealed strong gains on the concepts by pupils instructed by the interns. Several statistics confirm this result. The mean score derived from pretesting was 6.06. At the conclusion of the instructional task pupils were posttested and the mean score was found to be 10.68. The highest possible score was 15. In addition to the calculation of the overall means, each pupil's own gain score was computed. The following figure illustrates the range and extent of the gains made.

(Insert Figure 2 about here)

Three measures describe the pupils' gain central tendency: mean = 5.16, median = 5.0, and mode = 4.0.

Observation and Inference Task

The results of the observation and inference task also revealed strong gains by pupils instructed by the interns. Of the 89 students participating, 70 students showed a positive gain on the posttest. Only five students

showed a loss on the results of the posttest. All but one had equalled or exceeded the posttest mean score on the pretest, thus regressing to the mean.

(Insert Figure 3 about here)

The pretest mean for the pupils was 2.96; following instruction the posttest mean was found to be 5.21. The highest score possible was eight (8). The mean, median and mode of the pupils' gain score was 2.24, 2.0, and 2.0, respectively.

All three pupil performance measures provide evidence that pupils can demonstrate strong gains in concepts and skills when instructed in a variety of topics by field-based junior interns.

DISCUSSION

At this time, the intention of the teaching performance tasks (pupil performance measures) is not to evaluate individuals but rather to provide additional data on program effectiveness. It has been possible, however, to identify individual preservice teachers who excel or who do poorly. Popham (1973), also cautioned that "the recency of serious research attention to teaching performance tests as an evaluative tool probably renders them inappropriate at the present time for assessing individuals, other than those at the extremes of a distribution. . ." Even then, judgements should only be made when performance in a variety of tasks is consistent.

From the results of the "classification" pupil performance measure it can be inferred that early and intensive field experience as part of a teacher education program can increase the effectiveness of preservice teachers. In all fairness, it can also be argued that while statistically significant, the difference between groups in the attainment of one of the seven classification objectives may be viewed as an insignificant cognitive difference.

The gains of pupils on all three pupil performance measures can be inferred to be the result of the treatment (i.e., the instruction provided by the interns and fundamentally, their instruction as provided in the teacher education program). However, few verified assumptions beyond this can be made. While we hope that pupil gains can be attributed to the effectiveness of the teacher education program, there may be a number of other factors involved. Among these are:

1. sensitization to the format and style of the pretest may be responsible for improved performance on the posttest,

2. intervening variables, which are not identified, occurring between pre- and posttests and extraneous to the treatment, may also account for a portion of gain,

3. context variables--characteristics of the school environment about which interns and staff can do very little may have influenced the results,

4. random selection of pupils could not be assured due to the desires of some cooperating teachers,

5. except for the classification measure, interns administered both pre- and post-measures. It is our experience that many interns do not wish their pupils to "fail", even on the pretest. This leads some interns to "instruct" during the test. Recently, the adaptation of testing procedures had reduced this phenomena to a minimum.

These same factors are not peculiar to this study since they delimit the attempts of research to measure teacher effectiveness through the performance of children in classroom settings.

The question of validity of the pupil performance measures is of particular concern. While the effectiveness of teacher education programs and individual teachers can be inferred from pupil gains, it could also be inferred 1. that interns could produce similar gains as entry behavior and 2. that non-teachers (see Popham, 1971) could produce similar gains in pupils.

These issues have been raised in the hopes that our experience will be helpful to others who are considering the use of pupil performance

measures. In spite of a number of problems, the investigators believe that such measures have a place in a multi-dimensional research design for evaluating teacher and program effectiveness.

As well as being used as an effective evaluative technique pupil performance measures have proven highly effective as an instructional component of the teacher education program since they:

1. provide a context for systematic instruction, including the use of specified objectives, diagnostic and evaluative assessment and sequential planning,
2. provide an effective feedback mechanism focused on pupil growth
3. allow interns to rely on data rather than intuition on their relative effectiveness,
4. provide a common basis for interns to work cooperatively and share instructional ideas as well as comparative information on the cognitive learning styles of different age children.

Thus, pupil performance measures can be used in a formative sense to assist educators in understanding and improving instructional processes.

TABLE 1

MEANS AND STANDARD DEVIATIONS: CONSEQUENCE COMPONENT

STUDENT PERFORMANCE

GROUP	N	MEAN		ADJUSTED M OF THE DEPEND. VAR.	STANDARD DEVIATIONS	
		PRE	POST		PRE	POST
A (FIELD-BASED)	81	3.691	5.642	5.819	2.095	1.970
B (UNIVERSITY-BASED, FIELD ORIENTED)	121	4.182	5.008	4.890	2.137	2.027
TOTAL	202	3.985	5.262		2.129	2.024

TABLE 2

ADJUSTED ANALYSIS OF VARIANCE WITH PRETEST AS COVARIATE: CONSEQUENCE COMPONENT

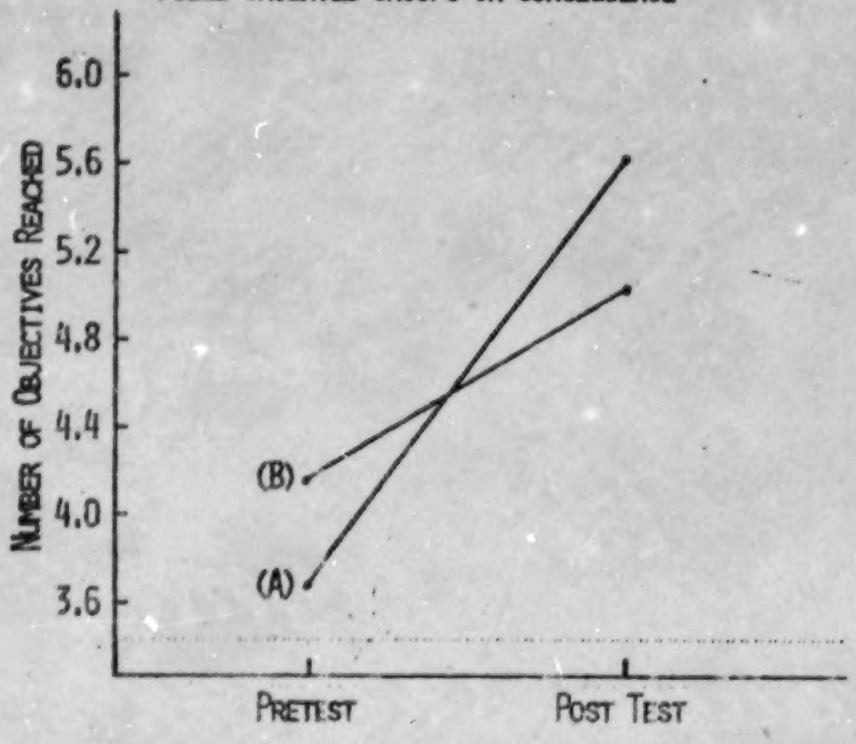
STUDENT PERFORMANCE

VARIABLE	SOURCE	DF	MS	F
LEARNING	TREATMENT	1	DATA WAS ANALYZED USING A COMPUTER AND THE WILKS LAMDA FORMULA WHICH DOES NOT YIELD A CONVENTIONAL MS.	17.182**
	ERROR	199		

**p<.01.

FIGURE 1

COMPARISON BETWEEN FIELD-BASED AND UNIVERSITY-BASED FIELD ORIENTED GROUPS ON CONSEQUENCE

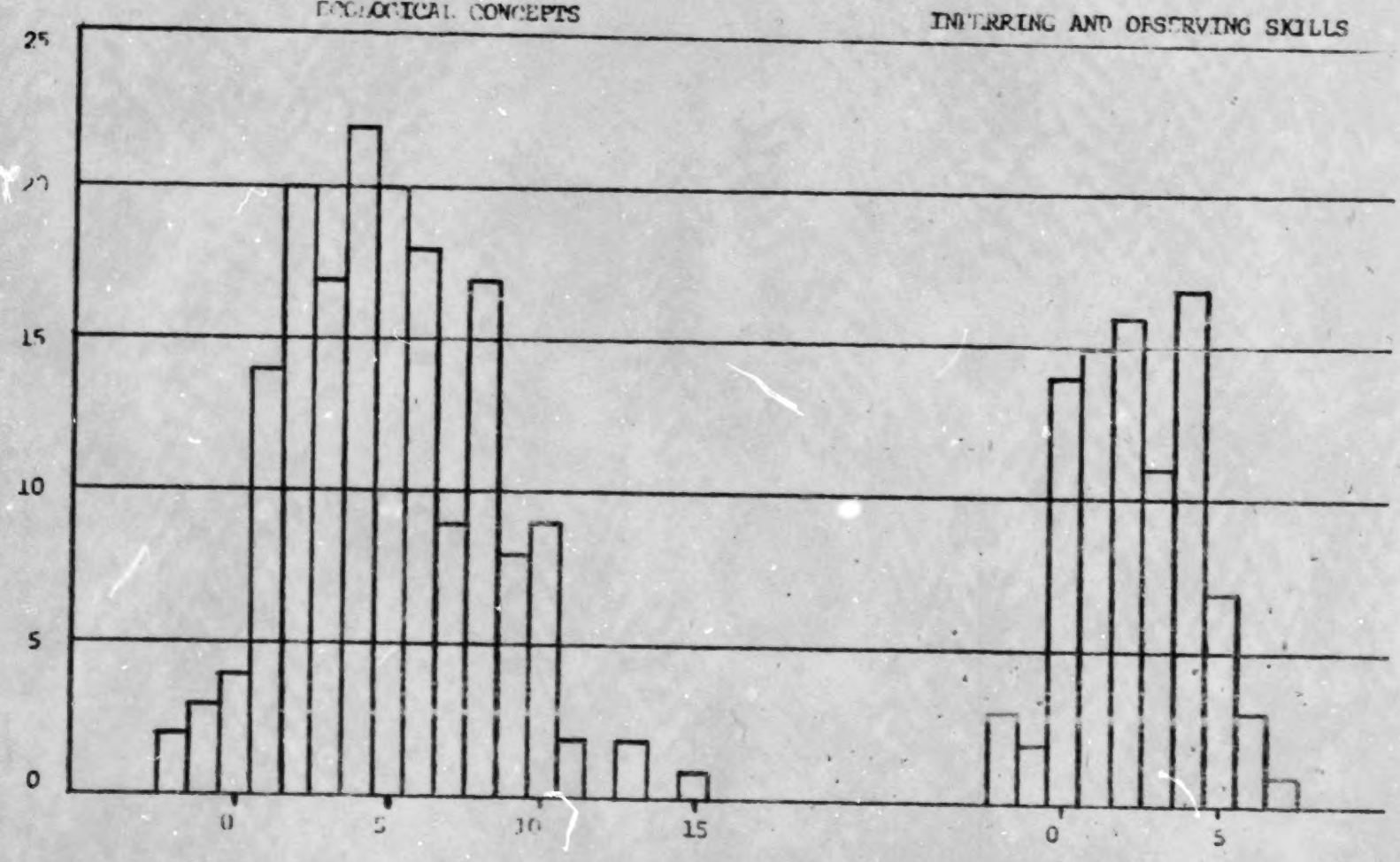


A = FIELD-BASED

B = UNIVERSITY-BASED, FIELD ORIENTED

FIGURE 2

PUPIL PERFORMANCE TESTS



NUMBER OF PUPILS

References

Cooper, J.M. Pupil growth measures and teacher evaluation: Pressing for practical applications. Paper presented at the meeting of the American Educational Research Association, Chicago, April, 1974.

Dunkin, M.J. and Biddle, M.J. The study of teaching. New York: Holt, Rinehart and Winston, Inc., 1974.

Hastie, J.C. The effect of a field-based teacher education program upon pupil learning. Paper presented at the 1973 annual meeting of the American Educational Research Association, New Orleans, February, 1973.

Kurtiz, T.B. A reliable measure of teacher effectiveness Educational Leadership, October, 1969.

Lahnston, A., Lapp, D., Rezba, R., and Willcutt, R. An investigation of the effects of training site on selected teacher training variables, a paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

Lapp, D., Lahnston, A., and Rezba, R. Final report: Exemplary Preservice Teacher Training Program Development Project: Right to Read Program Office of Education Project No. A00142AK, Grant No. OEG-0-74-9010, Boston University, Boston, MA July, 1975.

Lapp, D., Lahnston, A., and Rezba, R. Final report: Exemplary Preservice Teacher Training Program Development Project: Right to Read Program Office of Education Project No. A00142AK, Grant No. OEG-0-74-9010, Boston University, Boston, MA July, 1976 (in progress).

McNeil, J.D. and Popham, W.J. The assessment of teacher competence. In R.M. Travers (Ed.), Second Handbook of Research in Teaching. Chicago: Rand McNally, 1973.

Popham, W.J. Performance tests of teaching proficiency: Rationale, development, and validation. American Educational Research Journal 8 (1), January, 1971.

Popham, W.J. Identification and assessment of minimal competencies for objectives-oriented teacher education programs. Paper presented at the meeting of the American Educational Research Association, New Orleans, February, 1973.

Rezba, R., Lahnston, A., Lapp, D., and Willicutt, R. Pupil growth in classification skills as a consequence measure of learning site on preservice elementary teachers. A paper presented at the meeting of the National Association for Research in Science Teaching, Chicago, April, 1974.

Rosenshine, B. and Furst, N. Research in teacher performance criteria. In B.O. Smith (Ed.), Research in teacher education. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Schalock, H.D. The focus of performance based certification: Knowledge, teaching behavior, or the products that derive from a teacher's behavior. Paper presented for the Performance Based Certification Conference, Florida State Department of Education May, 1970. (ERIC Document Reproduction Service Number ED 045 550).

Smithman, H.H. and Lucio, Wm. H. Supervision by objectives: Pupil achievement as a measure of teacher performance. Educational Leadership, January, 1974.